Two Main Contents in a Syllabus for Language Documentation

the Learning Data Models and an Assignation of Data Conversion

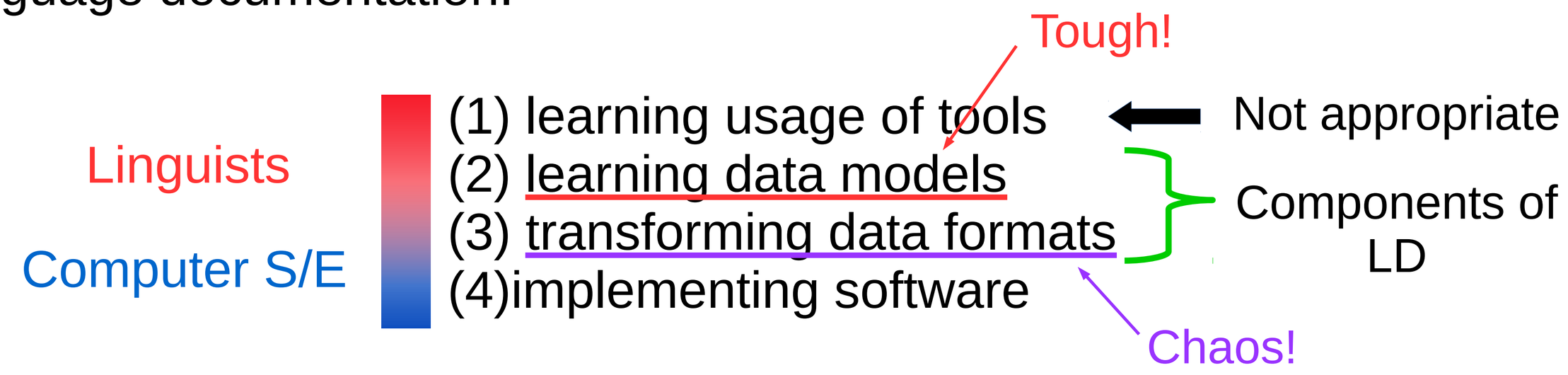
Kazushi Ohya
Tsurumi University

1. Components of Language Documentation

Four fields in language documentation:

(1) learning usages of tools including devices and software,
(2) learning data models that are used in programming based on computer science,
(3) transforming data formats of language data into others in moving to the next phase of data handling in language documentation, and
(4) implementing software of data management system such as ELAN, FLEx, SQL engines, web systems, and so on.

The components of language documentation are different according to the projects' scale, period, and members.
However, if we suppose linguists and computer people are participants of language documentation, we can get a clear perspective on a structure of language documentation.



2. Data Models

As far as a history of computer science and archive studies prove, there is no future in application dependence for long-term data preservation. The only solution is to learn data models developed based on computer science and to seek for a data model that is best for linguists themselves. However, there are not enough guidelines or textbooks on this topic.

3. Assignation of Data Conversion

In terms of projects' architecture, (c) is the most problematic point because there is a missing theory of data conversion: no format to define a pattern of a final description or converted data in a declarative form. Since no theory of data conversion, the heuristic solution is the only way for that. The software engineering provides a field of knowledge that we need to deal with this problem.